

Author's Accepted Manuscript

Rum classification using fingerprinting analysis of volatile fraction by headspace solid phase microextraction coupled to gas chromatography-mass spectrometry

José Raúl Belmonte-Sánchez, Simona Gherghel, Javier Arrebola-Liébanas, Roberto Romero González, José Luis Martínez Vidal, Ivan Parkin, Antonia Garrido Frenich



www.elsevier.com/locate/talanta

PII: S0039-9140(18)30492-2
DOI: <https://doi.org/10.1016/j.talanta.2018.05.025>
Reference: TAL18665

To appear in: *Talanta*

Received date: 22 March 2018
Revised date: 4 May 2018
Accepted date: 7 May 2018

Cite this article as: José Raúl Belmonte-Sánchez, Simona Gherghel, Javier Arrebola-Liébanas, Roberto Romero González, José Luis Martínez Vidal, Ivan Parkin and Antonia Garrido Frenich, Rum classification using fingerprinting analysis of volatile fraction by headspace solid phase microextraction coupled to gas chromatography-mass spectrometry, *Talanta*, <https://doi.org/10.1016/j.talanta.2018.05.025>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Rum classification using fingerprinting analysis of volatile fraction by headspace solid phase microextraction coupled to gas chromatography-mass spectrometry

José Raúl Belmonte-Sánchez¹, Simona Gherghel^{1,2,3}, Javier Arrebola-Liébanas¹, Roberto Romero González¹, José Luis Martínez Vidal¹, Ivan Parkin³, Antonia Garrido Frenich^{1*}

¹Research Group “Analytical Chemistry of Contaminants”, Department of Chemistry and Physics, *Research Centre for Agricultural and Food Biotechnology (BITAL)*, University of Almería, *Agrifood Campus of International Excellence, ceiA3*, E-04120 Almería, Spain.

²UCL Department of Security and Crime Science, 35 Tavistock Square, London, WC1H 9EZ, United Kingdom;

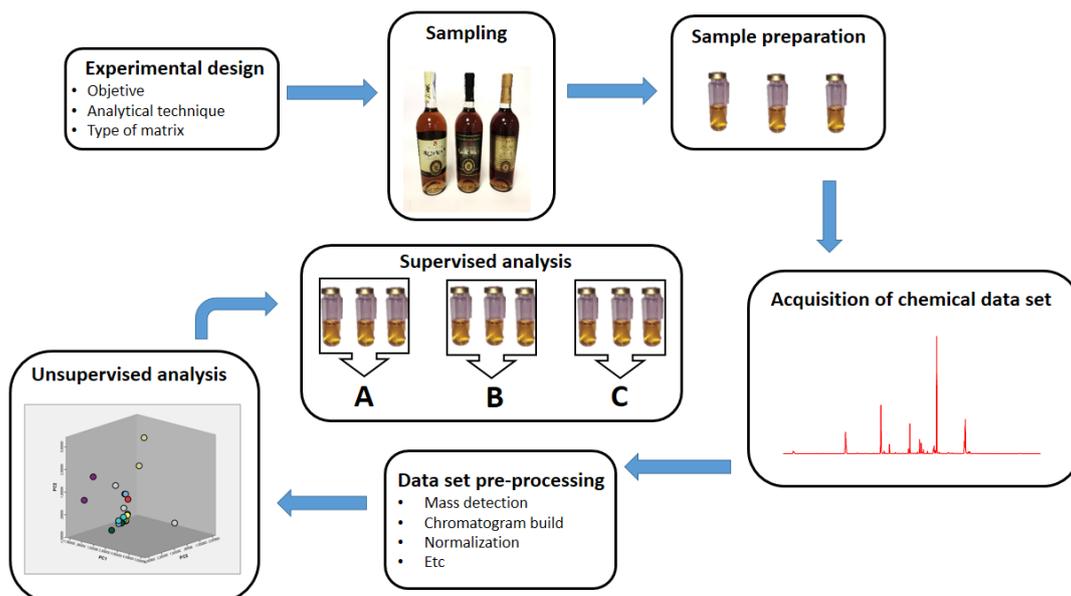
³UCL Department of Chemistry, 20 Gordon Street, London, WC1H 0AJ, United Kingdom.

* Corresponding author. Tel: +0034950015985; fax: +003495005008. E-mail address: agarrido@ual.es (A. Garrido Frenich).

Abstract

In this study, targeted and untargeted analyses based on headspace solid phase microextraction coupled to gas chromatography-mass spectrometry (HS-SPME-GC-MS) method were developed for classifying 33 different commercial rums. Targeted analysis showed correlation of ethyl acetate and ethyl esters of carboxylic acids with aging when rums of the same brand were studied, but presented certain limitations when the comparison was carried out between different brands. To overcome these limitations, untargeted strategies based on unsupervised treatments, such as hierarchical cluster analysis (HCA) and principal component analysis (PCA), as well as supervised methods, such as linear discriminant analysis (LDA) were applied. HCA allowed distinguishing main groups (with and without additives), while the PCA method indicated 40 ions corresponding to 13 discriminant compounds as relevant chemical descriptors for the correct rum classification (PCA variance of 88%). The compounds were confirmed based on the combination of retention indexes and low and high-resolution mass spectrometry (HRMS). Using the obtained results, LDA was carried out for the analytical discrimination of the remaining rums based on manufacturing country, raw material type, distillation method, wood barrel type and aging period and 94%, 91%, 92%, 95% and 94% of rums, respectively, were correctly classified. The proposed methodology has led to a robust analytical strategy for the classification of rums as a function of different parameters depending on the rum production process.

Graphical abstract



Keywords: rum, classification, volatile organic compounds, multivariate analysis, SPME-GC-MS

1. Introduction

Rum is a fairly aromatic spirit, obtained exclusively from sugar cane juice or molasses, and then subjected to the processes of alcoholic fermentation, distillation and aging. This spirit represents a widely popular alcoholic beverage with a high world consumption rate (more than 1 billion of litres per year) and an expected increase of 1.9% in volume terms over 2016-2021. [1,2].

The complex elaboration of this type of alcoholic beverage makes it an attractive object of study. Differences in the production process are known to lead to wide variability in its composition, although this variation has not been fully understood yet [3,4]. The production process begins with the fermentation of the chosen raw material, which leads to the formation of a number of volatile compounds, such as alcohols, ethyl esters and aldehydes, among others [5]. The resulting mash is distilled using heat in copper pot stills or in stainless steel columns to obtain a high content of ethanol, which inevitably leads to the loss of some aroma compounds [6,7]. Additionally, different distillation methods can be applied, such as continuous and batch distillation (e.g. Jamaican “heavy rums” typically made by batch distillation) [8]. The resulting distillate is diluted with pure demineralized water to obtain an alcohol percentage of around 35-40%, which is then aged in oak barrels previously used for whiskey or brandy production [9,10]. The aging step gives rum its characteristic flavor as a large number of new compounds emerge. Ethyl esters are generated as a result of the high percentage of ethanol, while a number of different compounds such as whiskey lactone, vanillin and 2-methoxyphenol can form because of the interaction with the wood barrels [3]. Additionally, as rum matures, it generally gains golden hues as a result of the tannins from the barrel staves [11]. After an aging period, typically of at least 1 year, the containers are opened for an optional blending step, where rums of different ages are

mixed to obtain specific organoleptic characteristics. Lastly, the colour and flavor of rums can be further modified by adding colorants and flavorings. Therefore, rums can be classified depending on the raw material, fermentation process, distillation process, aging period, type of barrel used, blending technique, alcohol strength and possible addition of additives.

Because of a lack of clear legislation around labelling, terms loosely related to aging periods, such as “Añejo”, “Dorado”, “Premium”, “Super Premium” or “Reserve” are often used by rum manufacturers without an actual quantitative/qualitative justification. Moreover, the age statement on labels is often not representative of the actual age, as blending of rums of different ages is carried out. According to legislation from both the European Union and the United States, the age statement on the label needs to refer to the youngest rum in the bottle [12,13]. However, in other countries, such as Canada, it can refer to the oldest rum [14]. Therefore, the development of methods that allow the reliable characterization of rums and an increased confidence of the consumers in this type of products in terms of authenticity is needed.

Nowadays, numerous methods have been described for the classification of alcoholic beverages based on the analysis of the volatile composition [9]. For that purpose, gas chromatography (GC) coupled to mass spectrometry (MS) has been one of the most frequently used technique [9]. In recent years, headspace solid phase microextraction (HS-SPME) has become the extraction method of choice. The combination of HS-SPME and GC-MS has been applied to different matrices such as wine [15–19], beer [20–24], tea beers [25] and other popular spirit beverages, such as whiskey [26,27], gin [28], or cocktail bitters [29].

However, to our knowledge, rum studies are less frequent and they have been generally limited to the comparison of this type of spirit with their South American

analogue (cachaça) [30], to ascertain a specific geographic origin (Cuban rums from non-Cuban rums) [31] or to the identification of some aroma indicators [6,32–35]. Due to the complexity and variability of rum preparation, their classification represents an analytical challenge.

To overcome this, multivariate analysis has been commonly employed for other such complex matrices in order to take advantage of the huge amount of data obtained from the GC-MS analysis. Unsupervised chemometric techniques as principal component analysis (PCA) [36,37] as well as hierarchical cluster analysis (HCA) [38,39] have been commonly used for a preliminary inspection of the data. Further supervised classification methods, such as linear discriminant analysis (LDA) [40,41] have been successfully applied for chemometric analysis, as well as for the classification of different types of beverages or foods [42].

The aim of this study has been the classification of various types of rums by developing a comprehensive and robust analytical strategy for the analysis of the volatile/semi-volatile compounds. After simple and completely automated HS-SPME-GC-MS analyses, the raw data were processed applying available statistical tools for targeted and untargeted analysis. For exploratory data analysis, unsupervised chemometric techniques using unlabelled data were applied. Afterwards, supervised techniques were applied to achieve rums classification based on the chemical correlations between samples.

2. Materials and methods

2.1 Reagents

Ethanol HPLC grade was obtained from J.T. Baker (Deventer, Holland). C7-C40 saturated alkanes standard mix (1000 $\mu\text{g}/\text{mL}$ in *n*-hexane) were supplied by Supelco (Bellefonte, PA, USA).

2.2 Samples

For this study, a total of 33 commercial rums were purchased from different local liquor stores (Almería, Spain). The rums were manufactured in 10 different countries: Cuba (5 samples), Dominican Republic (8 samples), Grenade (1 sample), Guatemala (3 samples), Jamaica (2 samples), Nicaragua (3 samples), Republic of Mauritius (2 samples), Spain (6 samples), Trinidad & Tobago (1 sample) and Venezuela (2 samples). All samples were stored in a refrigerator (4 °C) prior to analysis, in their original glass bottles. Information about the rum production from the official website of rum manufacturers as well as from the label, and assigned codes for each rum are summarized in Table 1. It should be pointed out that information about aging, raw material and distillation process was not provided by all manufacturers. When the information was not available, this was recorded as NA.

2.3 Sample preparation and HS-SPME procedure

Prior to HS-SPME-GC-MS analysis, the rum bottles were left to reach room temperature for 1 h. After that, they were opened for the first time. Rums with common origin were analysed equally across the sampling sequence according to a block design in order to guarantee their comparability and lack of potential analytical bias. Three replicates of each bottle were analysed.

Blanks which consisted of a mixture of Milli-Q water (J.T. Baker) and ethanol (Sigma-Aldrich; San Louis, MO, USA) at a ratio of 63:37 v/v were prepared to simulate

the alcohol content in a typical commercial rum. Blanks were analysed between each brand for various specific purposes: (i) to check the potential contamination generated by the septum (blank correction during the statistical treatment of the data), (ii) to evaluate potential carry over effect in the fiber and (iii) for additional cleaning up of the fiber.

For SPME extraction, different combinations of the selected parameters that are known to affect the fiber performance (sample volume, incubation time, extraction temperature, extraction time, and stirring speed) were applied in order to maximize the number and the intensity of volatile compounds extracted. Finally, ten mL of each rum sample were placed into a 20-mL glass vial fitted with a magnetic cap and a PTFE/silicone septum of 1.5 mm thickness. After 5 min of preheating the sample at 65 °C (continuous stirring, 250 rpm), the SPME fiber was exposed to the sample headspace for an adsorption time of 30 min with constant stirring (250 rpm).

After extraction, the fiber was inserted into the GC injector using a 0.8 mm dedicated SPME liner to allow thermal desorption of the analytes at a temperature of 250 °C for 2 min. The compounds were desorbed into the injector in splitless mode for 2 min, prior to the GC-MS analysis. After desorption, a fiber cleaning step was carried out for 6 additional min with an increased split rate of 100:1.

2.4 GC-QqQ-MS analysis

A Scion GC system equipped with an autosampler (Bruker Corporation, Freemont, CA, USA) was used for chromatographic analyses. Polydimethylsiloxane (100 µm film thickness) SPME fibers were obtained from Supelco (Bellefonte, Pennsylvania, USA). After their conditioning following manufacturer's recommendations, the fibers were used without any further modification. A VF-5ms

capillary column (30 m x 0.25 mm i.d. x 0.25 μ m film thickness) from Varian (Palo Alto, California, USA) was utilized for GC separation. Helium was used as carrier gas at a constant flow rate of 1 mL/min (36.7 cm/s linear velocity). An untreated fused silica capillary column (2 m x 0.25 mm) from Supelco was used as pre-column.

Mass spectrometric detection was performed by a triple quadrupole Scion QqQ-MS/MS (Bruker) operating in electron ionization mode (EI, 70 eV). Mass spectral data of the total ion chromatograms (TICs) and Kovats retention index (KI) of rum samples were compared to the NIST (2014) mass spectra database.

At the beginning of the analysis, the column temperature was set to 35 $^{\circ}$ C, and the temperature was increased to 100 $^{\circ}$ C at a 4 $^{\circ}$ C/min rate, and then to 250 $^{\circ}$ C (hold 20 min) at a rate of 20 $^{\circ}$ C/min. The total run time was 43.75 min.

The QqQ mass spectrometer was operated in full scan mode. The temperatures of the transfer line, manifold, and ionization source were set to 280, 40, and 280 $^{\circ}$ C, respectively. The electron multiplier voltage was set to 1600 V (+200 V offset above the value obtained in the auto-tuning process). Mass peak widths set in the first and third quadrupole were of m/z 1.5 and 2.0, respectively. The analysis was carried out in the range of mass/charge ratios of m/z 50-400.

The volatile compounds of interest were identified against a commercial library (NIST14) and by the use of GC retention indices. Retention time of each volatile was converted to the Kovats retention index using C7-C40 n-alkanes as references and verified with those reported in the literature. A retention index window of ± 20 was applied to MS peak identification assignment [43].

2.5 HRMS-Q-Exactive analysis

As an additional confirmation method, a Q-Exactive-GC hybrid quadrupole Orbitrap mass spectrometer (Q-Exactive™, Thermo Fisher Scientific, Bremen, Germany) was used with the same chromatographic conditions as the GC-QqQ-MS analysis reported above. A VF-5ms capillary column (30 m x 0.25 mm i.d. x 0.25 μm film thickness) from Varian (Palo Alto, California, USA) was used for GC separation. The positive electron ionization (EI) source was operated at 70 eV at a temperature of 200°C with a transfer line temperature of 250°C. High-resolution mode was operated at a 60000 full width at half maximum (FWHM) resolving power (m/z 207).

SPME extraction was operated using an identical fiber selected for GC-QqQ-MS analysis (polydimethylsiloxane, 100 μm film thickness). Quantitation was performed using Xcalibur 4.1 and TraceFinder 4.1 software (Thermo Fisher Scientific, Les Ulis, France).

2.5 Pre-processing and data treatment

The GC-MS raw data (XMS file format) were converted to CDF file format using the Openchrom software (Armonk, NY, USA) [44]. Then, the dataset was pre-processed with MZmine 2.23 software (Norwood, MA, USA) [45]. The MZmine's "3D Viewer" tool offered a three-dimensional representation of the total ion chromatogram (TIC) in order to work with useful ranges of intensity, retention time and m/z ranges, (**Fig. S-1, Supporting Information**). The pre-processing treatment consisted of the optimisation of the following steps: (i) centroid mass detection, (ii) chromatogram builder, (iii) chromatographic deconvolution, (iv) alignment, (v) peak list, (vi) duplicate peak filter and (vii) gap filling. The results were stored as a CSV format file.

The data set was imported as a XML file using Excel, version 2013 (Redmond, WA, USA) and then, it was processed by normalising the ion intensities between 0 and 1: $[\text{value} - \text{minimum value}] / [\text{maximum value} - \text{minimum value}]$.

2.6 Processing and multivariate data analysis

All statistical analyses in this study were conducted using the SPSS 23.0 software package (SPSS Inc, Chicago, IL, USA). For sample classification and discrimination, HCA, PCA and LDA were applied.

3. Results and discussion

3.1 HS-SPME-GC-MS method

A HS-SPME procedure was developed to carry out the analysis of the volatile/semivolatile compounds in rums. For that, a typical commercial 7-years old rum was used.

The 100 μm polydimethylsiloxane (PDMS) SPME fiber was selected as it has been extensively used for extracting volatile/semi-volatile compounds with a wide polarity range [46], some applications in alcoholic drinks such as wine [47] and rum [6,35].

For the GC-QqQ-MS analysis, the starting oven temperature was set to 35 $^{\circ}\text{C}$ to allow the elution of the most volatile compounds. In order to avoid the chromatographic co-elution of high concentrations of low molecular weight alcohols, such as ethanol and minor volatile compounds, each sample was monitored in full scan mode in the m/z range 50-400.

3.2 Analysis of targeted compounds

Targeted analysis was focused on the determination of some selected compounds traditionally present in rum samples and already reported as relevant compounds for the organoleptic properties of such spirit drinks [3,4,6]. For this purpose, ethyl acetate, furfural, 5-hydroxymethylfurfural, 2,5-furandicarboxaldehyde, vanillin, 3-methyl-1-butanol, whiskey lactone and the ethyl esters of the majoritarian carboxylic acids between C8 and C16 (individual and as sum of all them) were monitored and found in most of the studied rum samples. Their identification was carried out by comparison of their mass spectra with NIST mass spectrum library and considering the retention index tabulated in bibliography [43].

Six of the studied rum brands were available at different ages. For them, the peak areas of various discriminant compounds were plotted to check the content variation among the samples. The data for ethyl acetate and the sum of the even ethyl esters between C8 and C16 are shown in Fig. 1. The data obtained for the individual ethyl esters are provided in the Supporting Information Fig. S2. It can be observed that with a few exceptions, there is a trend for an increased peak area with rum aging within the same brand. For example, ethyl acetate and the sum of the ethyl esters were found at higher amounts in older rums (I, J, C and Q samples). However, these compounds content was not satisfactorily correlated with aging between samples of different brands and therefore, they cannot be used as proper chemical indicators of aging in rums. For example, in both plots it can be observed that some rum samples (i.e. C_15Y) contained lower amounts of ethyl acetate or ethyl esters than younger rum samples of different brands (i.e., A_7Y, I_5Y, H_8Y and H_12Y).

An evaluation of the presence of the target compounds in the different rums did not provide a clear correlation between their presence in the studied samples and other factors such as origin, raw material, distillation type, barrel type, etc. Therefore, to

overcome the limitations observed in the targeted analysis, further chemometrical tools were evaluated for untargeted analysis.

3.2 Pre-processing and dataset treatment

In the first stage, the raw data export to CDF format was visualized using the *3D Viewer* tool of MZmine 2.23. Details of the MZmine steps and parameters are summarized in Table S-1 (Supporting Information).

Once the alignment was completed and before the “*Gap Filling*” process, the extraneous ions detected in all the samples (corresponding to SPME fiber, rubber septum, column bleeding, etc.) were substrated. For that purpose, a blank sample was analysed between each analysis of every type of commercial rum. This allowed the possibility of subtracting the ions detected by the software in these samples (known interfering masses), that could lead to errors in classification. These ions were mostly represented by cyclosiloxane signals from the SPME fiber (m/z 222, 296, 370), siloxanes from the sample vial septa (m/z 73, 207, 281), and phthalates from plasticizers contamination (m/z 149).

The resulting data from MZmine were exported as a CSV file containing a matrix giving the ion intensity values for each rum sample. In the time window from 1.20 to 30.00 min for a mass range m/z from 50 to 400, when 96 analyses (33 different rum samples with three replicates each, except 3 rum samples that were studied in duplicate) were performed, 231 variables (number of ions detected as the relevant ones by MZmine from the GC-MS data) were obtained. Thus, a total of 22176 data points were processed and used for multivariate analysis.

3.3 Unsupervised methods: Exploratory data analysis

An exploratory data analysis by HCA was carried out. The HCA was applied in order to reduce the dimensions of the data set by grouping relatively similar samples in one cluster and relatively dis-similar objects in another. Besides, the HCA enabled to check the repeatability of the analyses, detect anomalous values and discard any outlier replicates. At this step, four outlier replicates out of a total of 96 rum analyses were discarded. However, a minimum of two replicates per rum sample were always considered.

In this first step, the HCA study classified satisfactorily five samples containing additives such as honey, syrup or flavoring (labelled as vanilla and tangerine flavored rums). The dendrogram obtained is shown in Fig. 2. Their chromatographic profiles showed relevant differences regards to the rest of rum samples. Therefore, and after concluding about the easy discrimination of rum samples containing additives, these kind of samples were disregarded to facilitate further chemometric classification of the rest of the rums.

Applying the factor analysis method, where the data was submitted to an orthogonal rotation (*varimax rotation*), a total of 40 variables (m/z) were selected as the most discriminant ions (score values higher than 0.80) for further PCA. The factor analysis scores obtained for all ions are shown in Table S-2 (Supporting Information).

PCA using the selected variables was applied to transform the high-dimensional variables into a small number of orthogonal factors, principal components (PCs), whilst accounting for the largest variance. PCA three-dimensional representation (Fig. 3) provided 88% of the total variance explained on the three first PCs. In this plot, PC1, PC2 and PC3 account for 36%, 36% and 15% of the individual variance, respectively.

3.4 Identification of the most discriminant compounds

These most discriminant ions were initially assigned to 13 volatile organic compounds based on their retention indexes and mass spectrometric data using the NIST database (match factor higher than 750). Further identification studies by HRMS with Q-Exactive Orbitrap were carried out and their mass spectra evaluated by a high-resolution filtering (HRF) tool from the Tracefinder software. Initially, the HRF assigned tentative identification was based on the use of high-resolution mass spectra but using the traditional spectral matching at unit resolution. Later, all unique combinations of atoms from these candidate precursors are generated and matched to m/z peaks using narrow mass tolerances [48]. The HRF scores obtained for the candidates ranged from 99.50% to 100.00%, confirming 11 of the compounds previously assigned by GC-QqQ-MS analysis using retention index and NIST database search (Table 2). Although different identification methods were applied for the same compound, further discussion was necessary for final confirmation of the three hesitant results.

The initial confirmation of ethyl acetate presented certain doubts because of the monitored mass range (m/z 50-400). Its predominant m/z 43 (CH_3CO^+) was not observed in those experimental conditions. Therefore, a new m/z range was investigated (m/z 35-400) for this particular case obtaining an adequate library spectrum comparison (Fig. S3 in Supporting Information) and a 100% HRF score. However, it should be noted that the use of a lower mass range has caused significant interferences for high volatility compounds (retention time lower than 5 min) due to the presence of high concentrations of ethanol (Fig. S4 in Supporting Information).

Two of the studied chromatographic peaks (retention time 18.4 and 19.5 min) were initially identified by NIST as ionene (1,2,3,4-tetrahydro-1,1,6-trimethylnaphthalene). However, HRF study of the HRMS data provided more reliable

confirmation for the peak at 18.4 min (HRF score of 100.00%). The chromatographic peak at retention 19.5 min also presented a high HRF score (99.53%) and probably it is a ionene-derivative, but not included in the NIST library. Ionene has been reported as a pyrolysis degradation of carotenoid products [49] that have been reported as present in sugarcane [50]. Therefore, ionene and its related compound can be produced during molasses production.

For the identification of the compound at RT 11.8 min, further discussion was required. After GC-QqQ-MS analysis, NIST library search showed trans-2-tetrahydro-5-methyl-furanmethanol as the main tentative identification option. However, further HRMS Q-Exactive Orbitrap analysis identified tetrahydro-2H-pyran-2-methanol as the first choice. Both compounds present the same formula ($C_6H_{12}O_2$), exact mass (m/z 116.08318) and very similar mass spectrometric profile. The main difference between both spectra is the distribution of ions in the cluster (m/z 55-60) shown in Fig. 4. The cluster profile obtained in the experimental spectrum fits with the one shown for tetrahydro-2H-pyran-2-methanol. This compound also presented a high HRF score of 100%.

Table 2 shows the 13 selected compounds with their corresponding identified compounds, empirical formula, selected ions, retention time, KI value, HRF score and identification methods used. It can be readily observed that all ions from PC1 and PC2 belong to six ethyl ester compounds (hexadecanoic acid, ethyl ester; (E)-9-octadecenoic acid, ethyl ester; tetradecanoic acid, ethyl ester; octanoic acid, 3-methylbutyl ester; decanoic acid, ethyl ester; octanoic acid, ethyl ester). Ethyl ester compounds, formed by the reaction of ethanol with acyl-CoA, are well-known for playing an important role in the organoleptic properties of fermented beverages, due to their fruity aroma [51]. For PC3, high discrimination power was observed for low molecular weight compounds,

such as ethyl acetate, diethoxymethane and 1,1 diethoxybutane. Ethyl acetate provides fruity and brandy notes to beverages [52]. Franitza et al. has already reported the presence of 1,1-diethoxy-3-methyl- butane in two different rums, providing intense fruity aroma [53]. Additionally, ionene and tetrahydro- 2H-Pyran-2-methanol were also identified as highly discriminant in PC3.

3.5 Supervised analyses

After variables selection, LDA was applied in order to classify the rums according to different groups by maximizing the ratio of between-class variance and minimizing the ratio of within-class variance. Classification groups were selected based on the main information about the rum elaboration, typically provided by rum manufacturers.

The leave-one-out method [54] was used as cross-validation procedure to evaluate the prediction ability (Q^2) for each LDA model using the previously selected variables. For a good predictability, the difference between R^2 (coefficient of determination) and Q^2 value should not exceed 0.3 and poor robustness of the model is usually suspected when that happens [55,56]. The rums that had no production information available for the respective classifying group were not used to ensure an accurate prediction.

For the country group, the 94% of the rums with 10 different origins were correctly classified ($R^2= 0.94$) (Fig. S-5 in Supporting Information). When cross-validation was carried out, the LDA model had a prediction ability of $Q^2= 0.73$. This cross-validation value can be explained because the composition of rum is mostly dependent on the manufacturing process due to its complexity and variability, rather than the manufacturing country.

When raw material was used as classificatory criterion (Fig. 5.a), 97% of rums were correctly classified. The molasses group was clearly separated from the sugar cane group, while the group containing the rums where no information was available were placed close to the molasses group. The LDA data distribution reveals that the NA rums with no information about their production process could be elaborated from molasses. The cross-validation process, which used only the sugar cane group and the molasses group, provided a prediction ability $Q^2 = 0.91$ indicating that clear differences between these two groups provide a good prediction ability of this model.

When the classification was performed according to distillation method, where four categories were used (copper pot stills, stainless steel columns, a mixture, and NA) (Fig. S-5 in Supporting Information), 92% of the rums were correctly classified. The cross-validation showed a prediction ability of $Q^2 = 0.87$ of the rums. Rums made in copper pot stills were clearly differentiated from the rest. The other three categories were correctly classified in most of the cases, despite their proximity in the LDA representation. Copper pot stills are the traditional method to distil rums; however, these days, many manufactures prefer stainless steel columns for their increased efficiency. With steel columns, only one fraction is collected, whereas copper pot stills provide manufacturers with the possibility to combine the head, heart and tail fractions to obtain specific rums. This elaboration difference can explain the clear LDA distinction of the copper pot distillation data.

For barrel type (unspecified oak barrel, American oak barrel, French oak barrel, NA), 95% of rums were correctly classified rums (Fig. S-5 in Supporting Information). A cross-validation value of $Q^2 = 0.86$ was obtained. The graphical proximity of the oak and American oak barrels could be an indication that American oak is the typical wood used when the manufacturers do not specify the type of barrel utilized.

The last LDA representation (Fig 5.b) was based on rum aging (young, medium, old)., The rum aging parameter was created by considering rums aged less than 5 years as young, those aged between 5-10 years as medium, and rums aged for more than years as old. A 94% of correctly classified rums were obtained for aging. For cross-validation, $Q^2 = 0.65$ were obtained. This value can be explained by the blending process and the differences in legislation about age labelling.

4. Conclusions

The present study has developed an analytical strategy for the classification of rums depending on different steps from the production process using the most discriminant compounds of the volatile fraction. For that, the targeted and untargeted analysis of rums was evaluated using HS-SPME-GC-MS and chemometric tools. The target analysis found some chemical indicators (ethyl acetate and ethyl esters of carboxylic acids) that could be correlated with aging within the same brand, but presented clear limitations when they were used across different brands. No other correlations with other parameters were found. On the other hand, the untargeted analysis using chemometric tools led to the classification of the 33 rums of different brands and ages. For that, unsupervised (HCA, PCA) and supervised techniques (LDA) were employed. The HCA showed considerable differences between the traditional rums and the rums prepared by addition of honey, syrup and flavouring. This chemometric method offers the potential to clearly distinguish these rums with additives from the rest. For the correct classification of traditional rums, PCA provided 40 ions as relevant chemical descriptors corresponding to 13 discriminant compounds (e.g. hexadecanoic acid ethyl ester, octanoic acid ethyl ester, decanoic acid ethyl ester, ethyl acetate and 1,1-diethoxy-3-methyl-butane). For the confirmation of the compounds, a

strategy based on the combination of retention indexes, NIST database matching using low-resolution mass spectrometry and HRF scores using high-resolution spectra obtained by HRMS Q-Exactive Orbitrap was employed.

The 28 traditional rums were classified based on manufacturing country, raw material, distillation method, barrel type, and aging period with data classification values from $R^2=0.94$ to $R^2=0.97$ (prediction ability from $Q^2=0.65$ to $Q^2=0.91$). LDA results showed conclusive differences between the classification groups and the importance of enough representative sample for each group. The applied classification strategy allowed for a better understanding of rum composition, including for rums with not accurate label information. Moreover, this technique could be of interest to future investigations, for example, to other spirit beverages.

Conflict of interest

The authors declare no conflict of interest.

References

- [1] Authentic Caribbean Rum, Decade of rum, (2013). <https://www.acr-rum.com/assets/pdf/decade-of-rum-english-usa.pdf>.
- [2] Top 25 Countries by Rum Consumption Per Capita, (n.d.). <https://blog.euromonitor.com/2017/08/top-25-countries-by-rum-consumption-per-capita.html> (accessed March 20, 2018).
- [3] L. Franitza, M. Granvogl, P. Schieberle, Influence of the production process on the key aroma compounds of rum: from molasses to the spirit, *J. Agric. Food Chem.* 64 (2016) 9041–9053. doi:10.1021/acs.jafc.6b04046.
- [4] J.A. Pino, S. Tolle, R. Gök, P. Winterhalter, Characterisation of odour-active compounds in aged rum, *Food Chem.* 132 (2012) 1436–1441. doi:10.1016/j.foodchem.2011.11.133.
- [5] D.R. Cardoso, L.G. Andrade-Sobrinho, A.F. Leite-Neto, R. V. Reche, W.D. Isique, M.M.C. Ferreira, B.S. Lima-Neto, D.W. Franco, Comparison between cachaça and rum using pattern recognition methods, *J. Agric. Food Chem.* 52 (2004) 3429–3433. doi:10.1021/jf035262+.
- [6] J. Pino, M.P. Martí, M. Mestres, J. Pérez, O. Busto, J. Guasch, Headspace solid-phase microextraction of higher fatty acid ethyl esters in white rum aroma, *J. Chromatogr. A.* 954 (2002) 51–57. doi:10.1016/S0021-9673(02)00167-X.
- [7] D.A. Nicol, Rum, in: A.G.H. Lea, J. Piggott (Eds.), *Fermented Beverage Prod.*, Springer US, 2003: pp. 263–287.
- [8] J.E. Murtagh, Feedstocks, fermentation and distillation for production of heavy and light rums, *Alcohol Textb. A Ref. Beverage, Fuel Ind. Alcohol Ind.* (2007) 243–255.
- [9] P. Wiśniewska, M. Śliwińska, T. Dymerski, W. Wardencki, J. Namieśnik,

Application of Gas Chromatography to Analysis of Spirit-Based Alcoholic Beverages, *Crit. Rev. Anal. Chem.* 45 (2015) 201–225.

doi:10.1080/10408347.2014.904732.

- [10] D.W. Lachenmeier, R. Attig, W. Frank, C. Athanasakis, The use of ion chromatography to detect adulteration of vodka and rum, *Eur. Food Res. Technol.* 218 (2003) 105–110. doi:10.1007/s00217-003-0799-8.
- [11] P. Hernández-Orte, E. Franco, C.G. Huerta, J.M. García, M. Cabellos, J. Suberviola, I. Orriols, J. Cacho, Criteria to discriminate between wines aged in oak barrels and macerated with oak fragments, *Food Res. Int.* 57 (2014) 234–241. doi:10.1016/j.foodres.2014.01.044.
- [12] CFR - Code of Federal Regulations of the United States of America, Alcohol, Tobacco Products and Firearms, Washington DC, 1991.
- [13] European Parliament Council of the European Union, 110/2008 of the European parliament and of the council of 15 January 2008 on the definition, description, presentation, labelling and the protection of geographical indications of spirit drinks and repealing Council Regulation (EEC) No 1576/89 Off, 2008.
- [14] Canada Department of Justice, Food and Drug Regulations, (2009) C.R.C, c 870.
- [15] A.E. Springer, J. Riedl, S. Esslinger, T. Roth, M.A. Glomb, C. Fauhl-Hassek, Validated modeling for german white wine varietal authentication based on headspace solid-phase microextraction online coupled with gas chromatography mass spectrometry fingerprinting, *J. Agric. Food Chem.* 62 (2014) 6844–6851. doi:10.1021/jf502042c.
- [16] Z. Xiao, S. Liu, Y. Gu, N. Xu, Y. Shang, J. Zhu, Discrimination of Cherry Wines Based on Their Sensory Properties and Aromatic Fingerprinting using HS-SPME- GC- MS and Multivariate Analysis, *J. Food Sci.* 79 (2014) 284–294.

- [17] Y. Niu, X. Zhang, Z. Xiao, S. Song, K. Eric, C. Jia, H. Yu, J. Zhu, Characterization of odor-active compounds of various cherry wines by gas chromatography–mass spectrometry, gas chromatography–olfactometry and their correlation with sensory attributes, *J. Chromatogr. B.* 879 (2011) 2287–2293.
- [18] S. Grützmann, V. Caliarì, M. Sganzerla, H. Teixeira, Talanta Volatile composition of Merlot red wine and its contribution to the aroma : optimization and validation of analytical method, *Talanta.* 174 (2017) 752–766.
doi:10.1016/j.talanta.2017.06.074.
- [19] A.K. Hjelmeland, P.L. Wylie, S.E. Ebeler, Talanta A comparison of sorptive extraction techniques coupled to a new quantitative , sensitive , high throughput GC – MS / MS method for methoxypyrazine analysis in wine, *Talanta.* 148 (2016) 336–345. doi:10.1016/j.talanta.2015.10.086.
- [20] J. Jiao, N. Ding, T. Shi, X. Chai, P. Cong, Z. Zhu, Study of Chromatographic Fingerprint of the Flavor in Beer by HS-SPME-GC, *Anal. Lett.* 44 (2011) 648–655.
- [21] O. Pinho, I.M. Ferreira, L.H. Santos, Method optimization by solid-phase microextraction in combination with gas chromatography with mass spectrometry for analysis of beer volatile fraction, *J. Chromatogr. A.* 1121 (2006) 145–153.
- [22] F. Rodrigues, M. Caldeira, J. de S. Câmara, Development of a dynamic headspace solid-phase microextraction procedure coupled to GC–qMSD for evaluation the chemical profile in alcoholic beverages, *Anal. Chim. Acta.* 609 (2008) 82–104.
- [23] D. Saison, D.P. De Schutter, F. Delvaux, F.R. Delvaux, Optimisation of a complete method for the analysis of volatiles involved in the flavour stability of

- beer by solid-phase microextraction in combination with gas chromatography and mass spectrometry, *J. Chromatogr. A.* 1190 (2008) 342–349.
- [24] N. Moreira, S. Meireles, T. Brandão, P. Guedes, D. Pinho, Talanta Optimization of the HS-SPME – GC – IT / MS method using a central composite design for volatile carbonyl compounds determination in beers, *Talanta.* 117 (2013) 523–531. doi:10.1016/j.talanta.2013.09.027.
- [25] L. Rong, L.J. Peng, C.T. Ho, S.H. Yan, M. Meurens, Z.Z. Zhang, D.X. Li, X.C. Wan, G.H. Bao, X.L. Gao, T.J. Ling, Brewing and volatiles analysis of three tea beers indicate a potential interaction between tea components and lager yeast, *Food Chem.* 197 (2016) 161–167. doi:10.1016/j.foodchem.2015.10.088.
- [26] C.N. Rhodes, K. Heaton, I. Goodall, P.A. Brereton, Gas chromatography carbon isotope ratio mass spectrometry applied to the detection of neutral alcohol in Scotch whisky: an internal reference approach, *Food Chem.* 114 (2009) 697–701.
- [27] J.I. Cacho, N. Campillo, P. Viñas, M. Hernández-córdoba, Talanta Cloud point extraction and gas chromatography with direct microvial insert thermal desorption for the determination of haloanisoles in alcoholic beverages, *Talanta.* 160 (2016) 282–288. doi:10.1016/j.talanta.2016.07.029.
- [28] S. Vichi, M. Riu-Aumatell, M. Mora-Pons, J.M. Guadayol, S. Buxaderas, E. López-Tamames, HS-SPME coupled to GC/MS for quality control of Juniperus communis L. berries used for gin aromatization, *Food Chem.* 105 (2007) 1748–1754.
- [29] A.J. Johnson, H. Heymann, S.E. Ebeler, Volatile and sensory profiling of cocktail bitters, *Food Chem.* 179 (2015) 343–354. doi:10.1016/j.foodchem.2015.01.114.
- [30] E.S.P. Nascimento, D.R. Cardoso, D.W. Franco, Quantitative ester analysis in cachaça and distilled spirits by gas chromatography– mass spectrometry (GC–

- MS), *J. Agric. Food Chem.* 56 (2008) 5488–5493.
- [31] O.M. Sampaio, R.V. Reche, D.W. Franco, Chemical profile of rums as a function of their origin. The use of chemometric techniques for their identification, *J. Agric. Food Chem.* 56 (2008) 1661–1668. doi:10.1021/jf0726841.
- [32] F.W.B. Aquino, L.M. Boso, D.R. Cardoso, D.W. Franco, Amino acids profile of sugar cane spirit (cachaça), rum, and whisky, *Food Chem.* 108 (2008) 784–793.
- [33] Z.L. Cardeal, P.J. Marriott, Comprehensive two-dimensional gas chromatography-mass spectrometry analysis and comparison of volatile organic compounds in Brazilian cachaça and selected spirits, *Food Chem.* 112 (2009) 747–755. doi:10.1016/j.foodchem.2008.06.057.
- [34] E.A. Nonato, F. Carazza, F.C. Silva, C.R. Carvalho, Z. de L. Cardeal, A headspace solid-phase microextraction method for the determination of some secondary compounds of Brazilian sugar cane spirits by gas chromatography, *J. Agric. Food Chem.* 49 (2001) 3533–3539.
- [35] J.A. Pino, Characterization of rum using solid-phase microextraction with gas chromatography-mass spectrometry, *Food Chem.* 104 (2007) 421–428. doi:10.1016/j.foodchem.2006.09.031.
- [36] R.S. da Costa, S.R.B. Santos, L.F. Almeida, E.C.L. Nascimento, M.J.C. Pontes, R.A.C. Lima, S.S. Simões, M.C.U. Araújo, A novel strategy to verification of adulteration in alcoholic beverages based on Schlieren effect measurements and chemometric techniques, *Microchem. J.* 78 (2004) 27–33.
- [37] J. Bouhlef, D.J. Bouveresse, S. Abouelkaram, E. Baéza, C. Jondreville, A. Travel, J. Ratel, E. Engel, D.N. Rutledge, *Talanta* Comparison of common components analysis with principal components analysis and independent

- components analysis : Application to SPME-GC-MS volatolomic signatures, *Talanta*. 178 (2018) 854–863. doi:10.1016/j.talanta.2017.10.025.
- [38] U. Römisch, H. Jäger, X. Capron, S. Lanteri, M. Forina, J. Smeyers-Verbeke, Characterization and determination of the geographical origin of wines. Part III: multivariate discrimination and classification methods, *Eur. Food Res. Technol.* 230 (2009) 31–45.
- [39] A. Montaña, *Talanta Volatile profiles of green Spanish-style table olives* : Application of compositional data analysis for the segregation of their cultivars and production areas, *Talanta*. 169 (2017) 77–84.
doi:10.1016/j.talanta.2017.03.066.
- [40] A. Biancolillo, R. Bucci, A.L. Magrì, A.D. Magrì, F. Marini, Data-fusion for multiplatform characterization of an Italian craft beer aimed at its authentication, *Anal. Chim. Acta*. 820 (2014) 23–31.
- [41] J. Saurina, Characterization of wines using compositional profiles and chemometrics, *TrAC Trends Anal. Chem.* 29 (2010) 234–245.
- [42] F. Magagna, L. Valverde-Som, C. Ruíz-Sambl As, L. Cuadros-Rodríguez, S.E. Reichenbach, C. Bicchi, C. Cordero, Combined untargeted and targeted fingerprinting with comprehensive two-dimensional chromatography for volatiles and ripening indicators in olive oil, (2016). doi:10.1016/j.aca.2016.07.005.
- [43] V.I. Babushok, Trends in Analytical Chemistry Chromatographic retention indices in identification of chemical compounds, *Trends Anal. Chem.* 69 (2015) 98–104. doi:10.1016/j.trac.2015.04.001.
- [44] P. Wenig, J. Odermatt, OpenChrom: a cross-platform open source software for the mass spectrometric analysis of chromatographic data, *BMC Bioinformatics*. 11 (2010) 405. doi:10.1186/1471-2105-11-405.

- [45] T. Pluskal, S. Castillo, A. Villar-Briones, M. Orešič, MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, *BMC Bioinformatics*. 11 (2010) 395. doi:10.1186/1471-2105-11-395.
- [46] S. Risticevic, D. Vuckovic, J. Pawliszyn, Solid - Phase Microextraction DEVELOPMENT OF FIBER SOLID - PHASE, (2010).
- [47] A. Ruiz-Delgado, F.J. Arrebola-Liébanas, R. Romero-González, R. López-Ruiz, A. Garrido Frenich, Headspace solid-phase microextraction coupled to gas chromatography-tandem mass spectrometry for the determination of haloanisoles in sparkling (cava and cider) and non-sparkling (wine) alcoholic beverages, *Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess.* 33 (2016) 1535–1544. doi:10.1080/19440049.2016.1229870.
- [48] N.W. Kwiecien, D.J. Bailey, M.J.P. Rush, J.S. Cole, A. Ulbrich, A.S. Hebert, M.S. Westphall, J.J. Coon, High-Resolution Filtering for Improved Small Molecule Identification via GC/MS, *Anal. Chem.* 87 (2015) 8328–8335. doi:10.1021/acs.analchem.5b01503.
- [49] R.G. Achari, G. Shaw, R. Holleyhead, Identification of ionene and other carotenoid degradation products from the pyrolysis of sporopollenins derived from some pollen exines, a spore coal and the Green River shale, *Chem. Geol.* 12 (1973) 229–234. doi:10.1016/0009-2541(73)90091-0.
- [50] J.D. Fontana, B. Czczuga, T.M.B. Bonfim, M.B. Chociai, B.H. Oliveira, M.F. Guimarães, M. Baron, Bioproduction of carotenoids: The comparative use of raw sugarcane juice and depolymerized bagasse by *Phaffia rhodozyma*, *Bioresour. Technol.* 58 (1996) 121–125. doi:10.1016/S0960-8524(96)00092-2.
- [51] K.J. Verstrepen, G. Derdelinckx, J.P. Dufour, J. Winderickx, J.M. Thevelein, I.S.

- Pretorius, F.R. Delvaux, Flavor-active esters: Adding fruitiness to beer, *J. Biosci. Bioeng.* 96 (2003) 110–118. doi:10.1016/S1389-1723(03)90112-5.
- [52] A.P. Dionísio, G. Molina, D. Souza de Carvalho, R. Dos Santos, J.L. Bicas, G.M. Pastore, Natural flavourings from biotechnology for foods and beverages, in: *Nat. Food Addit. Ingredients Flavourings*, 2012: pp. 231–259. doi:10.1016/B978-1-84569-811-9.50011-6.
- [53] L. Franitza, M. Granvogl, P. Schieberle, Characterization of the Key Aroma Compounds in Two Commercial Rums by Means of the Sensomics Approach, *J. Agric. Food Chem.* 64 (2016) 637–645. doi:10.1021/acs.jafc.5b05426.
- [54] D.W. Osten, Selection of optimal regression models via cross-validation, *J. Chemom.* 2 (1988) 39–48. doi:10.1002/cem.1180020106.
- [55] R. Veerasamy, H. Rajak, A. Jain, S. Sivadasan, C.P. Varghese, R.K. Agrawal, Ravichandran Veerasamy, et al : Validation of QSAR Models-Strategies and Importance Validation of QSAR Models -Strategies and Importance, (n.d.). <https://pdfs.semanticscholar.org/4eb2/5ff5a87f2fd6789c5b9954eddddf1c59dab.pdf> (accessed January 24, 2018).
- [56] B. Worley, R. Powers, *HHS Public Access*, 1 (2015) 92–107. doi:10.2174/2213235X11301010092.Multivariate.

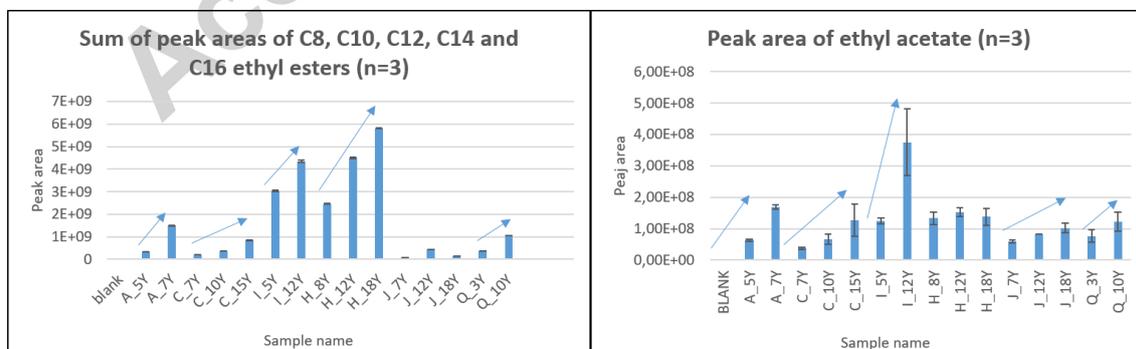
Figure 1 Bar chart representation of ethyl acetate (left) and the sum ethyl esters between C8 and C16 (right) among different rum samples. Arrows were added to highlight aging trends within the same brands.

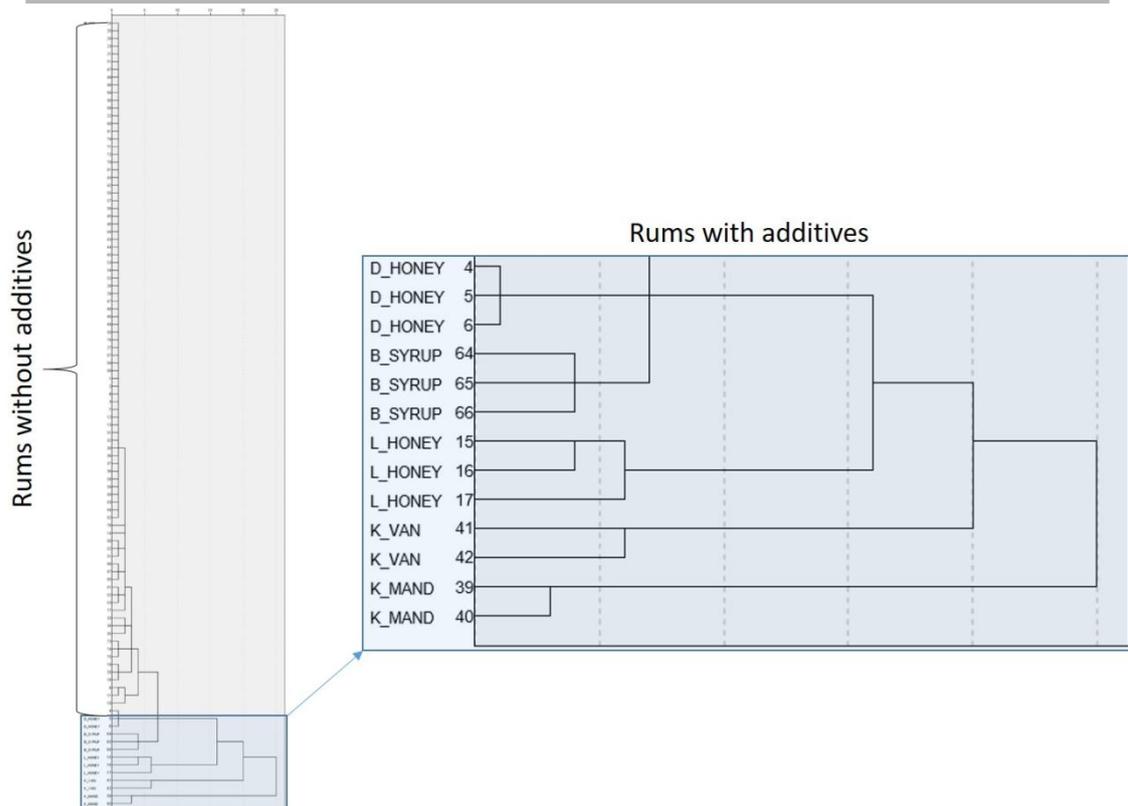
Figure 2. Dendrogram obtained by HCA using the total rum batch. See Table 1 for sample code.

Figure 3. PCA representation using the most discriminant ions.

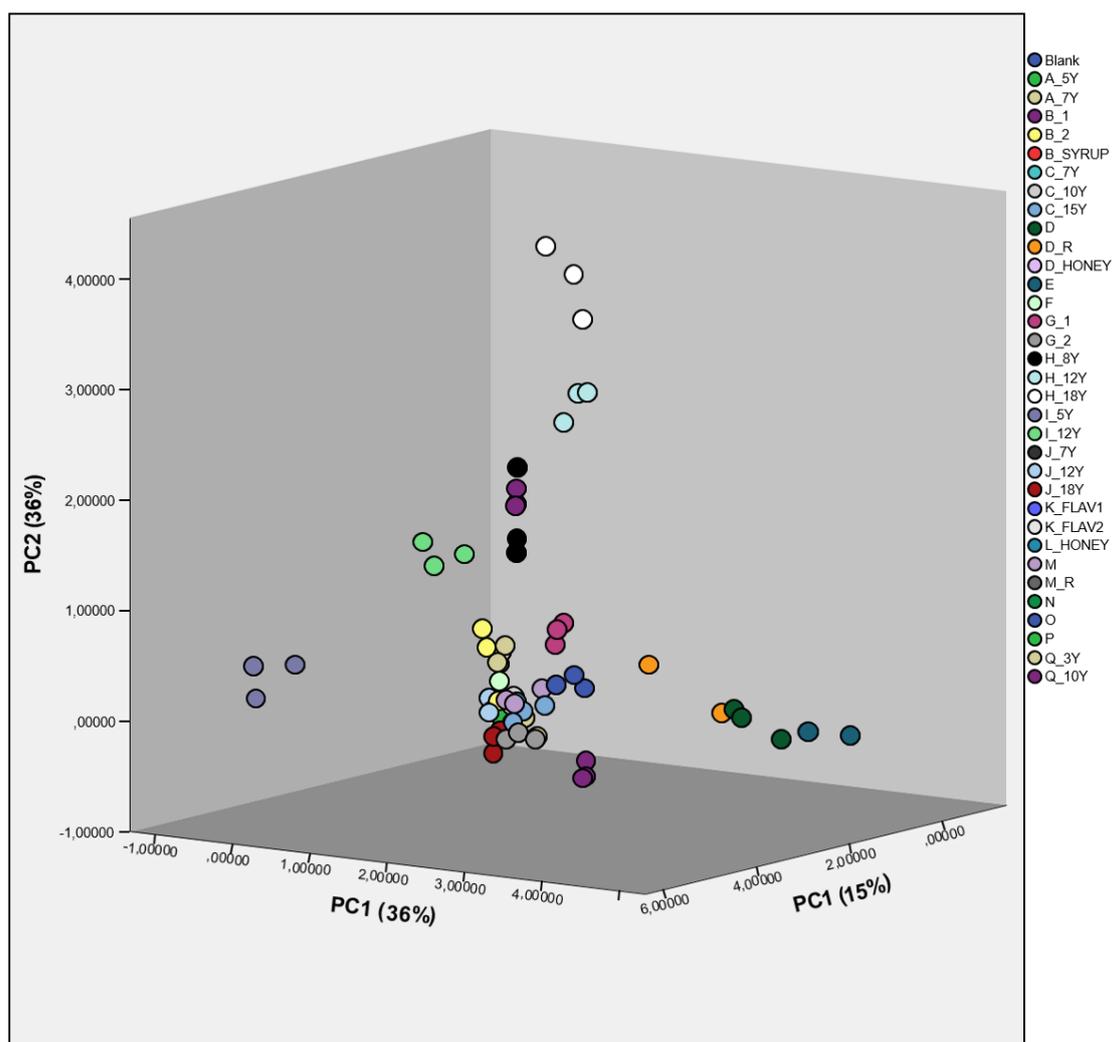
Figure 4 Comparison between theoretical mass spectrum of tetrahydro-2H-pyran-2-methanol (a), theoretical mass spectrum of trans-tetrahydro-5-methyl-2-furanmethanol (b) and the obtained experimental spectrum by HRMS Q-Exactive Orbitrap (c) with zoom-in on the m/z 55-60 cluster (d).

Figure 5. LDA representation showing the classification of 28 rums based on (a) raw material and (b) aging period.

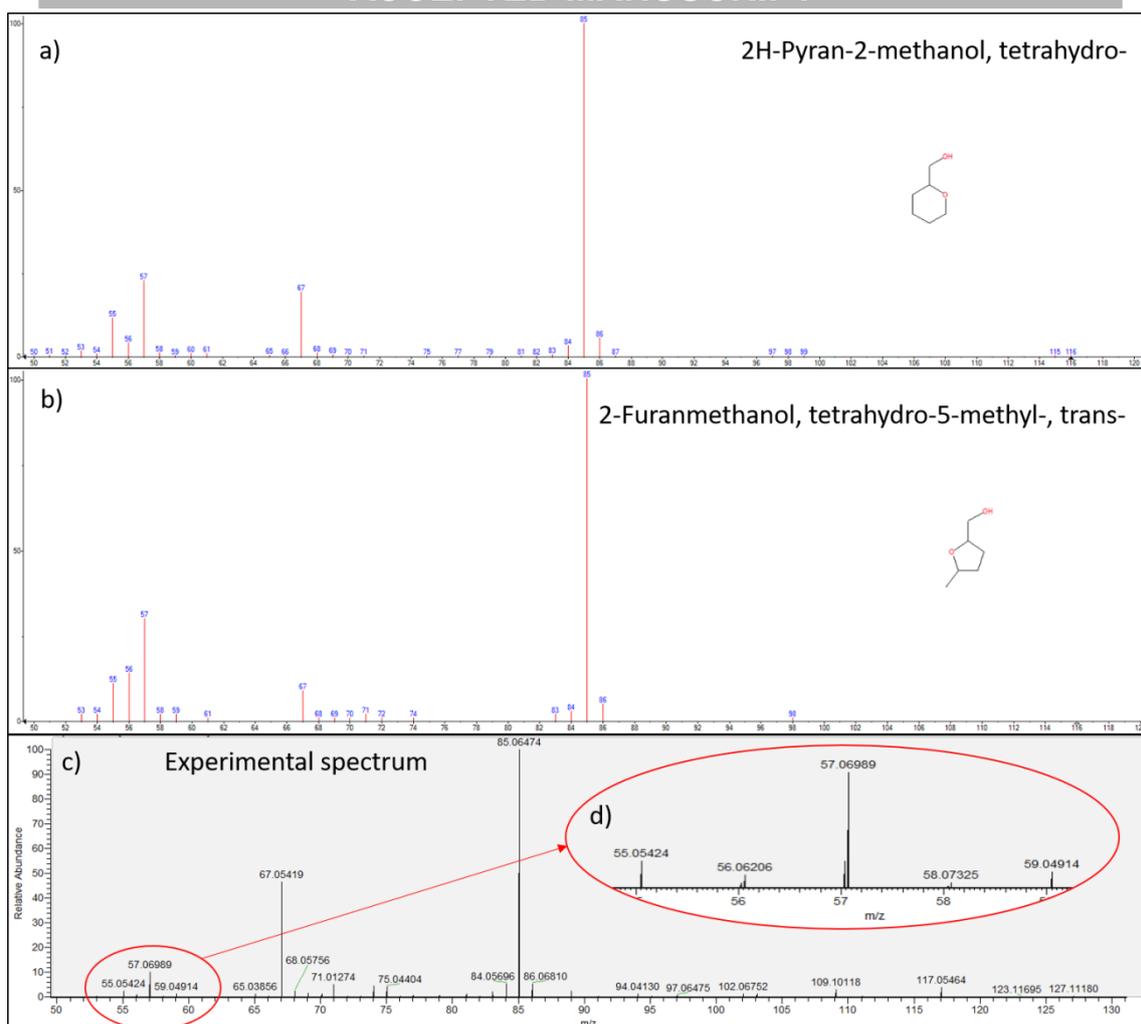


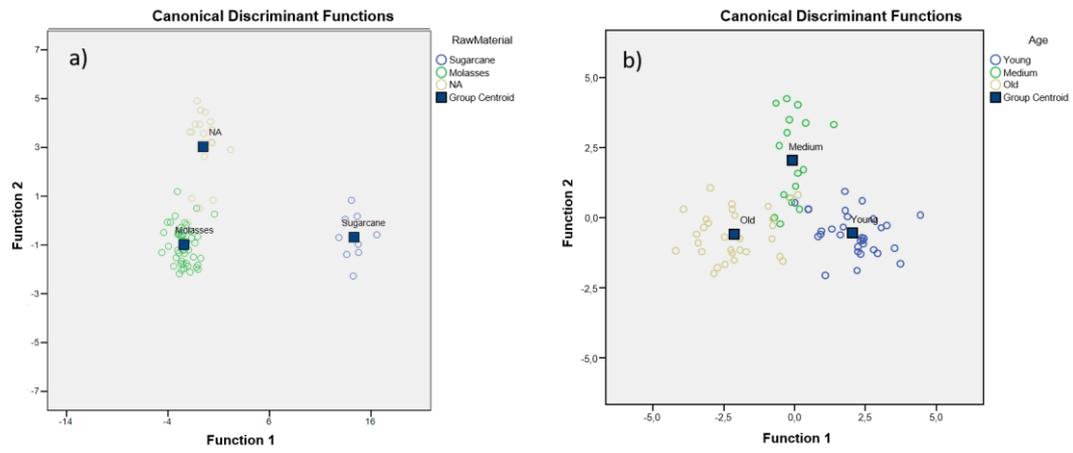


Accepted manuscript



Accepted





Accepted manuscript

Table 1. Code, bottle labelling and manufacturer information for the rum samples analysed

Brand code	Bottle code	Origin	Aging	Raw material	Distillation method	Barrel
A	A_5Y	Cuba	5 years	Molasses	Stainless steel columns	Oak barrel
	A_7Y		7 years	Molasses	Stainless steel columns	Oak barrel
B	B_1		Anejo	Molasses	N/A	American oak barrel
	B_2	Young	Molasses	N/A	American oak barrel	
	B_SYRUP1	N/A	Molasses	N/A	American oak barrel	
C	C_7Y	Dominican Republic	7 years	Molasses	Stainless steel columns	American oak barrel
	C_10Y		10 years	Molasses	Stainless steel columns	American oak barrel
	C_15Y		15 years	Molasses	Stainless steel columns	American oak barrel
D	D		Anejo	N/A	N/A	N/A
	D_R2		Anejo	N/A	N/A	N/A
	D_HONEY1		N/A	N/A	N/A	N/A
E	E		Old reserve	N/A	N/A	N/A
F	F	Anejo	N/A	N/A	N/A	
G	G_1	Grenada	Double aged.	Molasses	N/A	Mixed (American

			Old reserve			oak barrel+ French oak barrel)
	G_2	Trinidad & Tobago	Double aged. Overproof	Molasses	Stainless steel columns	Mixed (American oak barrel+ French oak barrel)
H	H_8Y	Guatemala	8 years	Sugarcane juice	Stainless steel columns	American oak barrel
	H_12Y		12 years	Sugarcane juice	Stainless steel columns	American oak barrel
	H_18Y		18 years	Sugarcane juice	Stainless steel columns	American oak barrel
I	I_5Y	Jamaica	5-10 years	Molasses	Copper pot stills	American oak barrel
	I_12Y		12 years	Molasses	Copper pot stills	American oak barrel
J	J_7Y	Nicaragua	7 years	Molasses	Stainless steel columns	American oak barrel
	J_12Y		12 years	Molasses	Stainless steel columns	American oak barrel
	J_18Y		18 years	Molasses	Stainless steel columns	American oak barrel
K	K_MAND1	Republic	Double	Sugarcane	Stainless	N/A

		of Mauritius	aged	juice	steel columns	
	K_VAN1		Double aged	Sugarcane juice	Stainless steel columns	N/A
L	L_HONEY1	Spain	N/A	N/A	N/A	N/A
M	M		Anejo	Molasses	N/A	N/A
	M_R2				Molasses	N/A
N	N		Dorado	N/A	N/A	N/A
P	O		Anejo	N/A	N/A	N/A
O	P		N/A	N/A	N/A	N/A
Q	Q_3Y	Venezuela	3 years	Molasses	Mixed (Steel column+a bit of copper)	American oak barrel
	Q_10Y		10 years	Molasses	Mixed (Steel column+a bit of copper)	French oak barrel

1: additives added

2: same rum type purchased at a different liquor store

N/A; information not available

Table 2. Most discriminant ions for PC1, PC2, and PC3 with their score, retention time and corresponding tentatively identified compounds

PC	Compound	Empirical formula	m/z (score)	RT (min)	KI ^a	HRF (%) ^b	Identification method
1	Hexadecanoic acid, ethyl ester	C ₁₈ H ₃₆ O ₂	185.2 (0.94); 241.2 (0.94); 284.3 (0.93); 213.2 (0.93); 157.2 (0.89); 115.1 (0.88); 88.1 (0.86); 55.1 (0.85);	24.10	1993	100.00	KI, Lib ^c , HRF
1	(E)-9-Octadecenoic acid ethyl ester	C ₂₀ H ₃₈ O ₂	180.2 (0.91); 112.2 (0.90); 222.3 (0.87); 264.2 (0.86); 137.1 (0.85); 109.1 (0.85); 169.1 (0.84)	25.17	2170	100.00	KI, Lib, HRF

1	Tetradecanoic acid, ethyl ester	$C_{16}H_{32}O_2$	88.1 (0.86); 55.1 (0.82)	23.05	1789	100.00	KI. Lib. HRF
2	Octanoic acid, 3-methylbutyl ester	$C_{13}H_{26}O_2$	70.2 (0.95); 127.2 (0.92)	20.89	1441	99.83	KI. Lib. HRF
2	Decanoic acid, ethyl ester	$C_{12}H_{24}O_2$	157.1 (0.92); 115.1 (0.92); 88.1 (0.81); 55.1 (0.91)	20.44	1389	99.50	KI. Lib. HRF
2	Octanoic acid, ethyl ester	$C_{10}H_{20}O_2$	88.1 (0.85); 127.2 (0.85); 57.2 (0.85)	18.17	1194	99.86	KI. Lib. HRF
3	Ionene	$C_{13}H_{18}$	131.1 (0.94); 159.1 (0.91); 144.1 (0.90); 116.1 (0.85); 91.1 (0.85); 113.1 (0.83)	18.4	1211	100.00	KI. Lib. HRF
3	Butane, 1,1-diethoxy-3-methyl-	$C_9H_{20}O_2$	103.1 (0.94); 73.1 (0.94); 75.1 (0.94)	9.2	918	99.82	KI. HRF
3	Ionene-derivative	$C_{13}H_{18}$	159.1 (0.90)	19.5	1295	99.53	HRF

3	Unknown	-	130.2 (0.89)	18.2	1197	-	-
3	Ethyl Acetate	C ₄ H ₈ O ₂	54.7 (0.87)	2.3	611	100.00	KI. Lib. HRF
3	2H-Pyran-2- methanol, tetrahydro-	C ₆ H ₁₂ O ₂	85.1 (0.86)	11.8	988	100.00	KI. Lib. HRF
3	Methane, 1.1- diethoxy-	C ₅ H ₁₂ O ₂	103.1 (0.84)	3.3	668	-	KI. Lib

^a KI; Kovats retention indices.

^b HRF (High-Resolution Filtering score); percentage of the spectrum obtained by HRMS Orbitrap that can be explained by combination of accurate mass, library matching and percentage of explained ions observed.

^c Lib; Identification based on mass spectrometric data using the NIST database (match factor higher than 750).

HIGHLIGHTS

- Untargeted analysis based on HS-SPME-GC-MS has been used for rum classification
- 33 different commercial rums from various brands and different ages were analyzed
- The most discriminant compounds of the volatile fraction of rums has been utilized
- Unsupervised and supervised treatments such as HCA, PCA and LDA were applied
- LDA provides suitable classification considering different factors as raw material

Accepted manuscript